Early detection of breast cancer using mammography images and software engineering process

Muayad Sadik Croock[1], Saja Dhyaa Khuder[2], Ayad Esho Korial[3], Sahar Salman Mahmood[4]
[1,2,3]Computer Engineering Department, University of Technology, Iraq
[4]Civil Engineering Department, University of Technology, Iraq

Article Info

ABSTRACT

The breast cancer has affected a wide region of women as a particular case. Therefore, different researchers have focused on the early detection of this disease to overcome it in efficient way. In this paper, an early breast cancer detection system has been proposed based on mammography images. The proposed system adopts deep-learning technique to increase the accuracy of detection. The convolutional neural network (CNN) model is considered for preparing the datasets of training and test. It is important to note that the software engineering process model has been adopted in constructing the proposed algorithm. This is to increase the reliably, flexibility and extendibility of the system. The user interfaces of the system are designed as a website used at country side general purpose (GP) health centers for early detection to the disease under lacking in specialist medical staff. The obtained results show the efficiency of the proposed system in terms of accuracy up to more than 90% and decrease the efforts of medical staff as well as helping the patients. As a conclusion, the proposed system can help patients by early detecting the breast cancer at far places from hospital and referring them to nearest specialist center.

Corresponding Author:

Muayad Sadik Croock,
Department of Computer Engineering,
University of Technology-Iraq,
Alsinaa Street, Baghdad, Iraq.
Email: Muayad.S.Croock@uotechnology.edu.iq

1. INTRODUCTION

Recently, the breast cancer is considered as the most dangerous risk that attacks the life of women. This disease is the result of different reasons, such as the life style and inheritance effects. The detection of this disease is based on allocating the changing in the soft tissue of the breast in early level. X-ray based mammography images is normally adopted for breast cancer detection. These images have been taken in different angles to cover all parts of the disease. It is well known that the X-ray images suffers from low contrasting due to low volume of radiation for health reason. Thus, different methods are used for implementing the image enhancement including artificial intelligent strategies and deep learning [1, 2]. The deep-learning technology has been considered in detecting of different diseases. In this work, we adopt the convolutional neural network (CNN) based deep-learning method for detecting the disease. It includes the pre-processing stage that enhance the considered images to increase the contrast and how the tissues if breast clearly. The CNN is used for constructing the model of extracting the features of included images. These features are

used to build the training dataset and detecting the disease of test samples. As a result, the processing time is reduced as well in efficient way due to low size of underlying images [2-5].

As mentioned earlier, the researchers interested in previous work about detecting different types of disease using deep-learning. It based on classifying these diseases into classes based on the features of underlying images. In [1], a new deep learning classifier has been proposed based on digital mammography images. The authors introduced a classifier for detecting the tumor tissues, in addition to overcoming the problem of the low contrast images. The contour function was used based on Chan-Vese level set method. Moreover, the required features were extracted using deep learning based CNN. the false results have been reduced by adding a complex valued relaxation to the classifier, while the accuracy is increased up to 99%. In [2], the authors presented a method of learning for a feature hierarchy of unlabeled dataset. The dataset was entered to the classifier for segmenting the breast density and scoring the mammography texture. Both of lifetime and population sparsity were considered in the proposed regularizer, used for controlling the extendibility of the presented model. This method was presented to ease the implementation and the obtained results ensured the high accuracy. In [3], the authors solved the problem of the risky development of this disease, appeared in the investigated images using cranio-caudal (CC) and mediolateral oblique (MLO). A deep learning model was used for tackling the problem of unregistered breast images and related segmentations. These parameters can affect the performance of the proposed method in bad way. The authors of [4], adopted different deep learning approaches for detecting and investigated of breast cancer using ultrasound session. The approaches of Patch-based LeNet, a U-Net, and a transfer learning in combination with a pertained FCN-AlexNet had been utilized for achieving the objective of the presented approaches.

The obtained results showed the high accuracy in comparison with the traditional methods. In [5], a tomosynthesis classification method was proposed using CNN based deep learning. More than 300 mammography images were collected from University of Kentucky. The utilized of deep learning was built to design a classifier for working on 2D and 3D images. The achieved results explained the superior performance of the proposed method. In [6], the authors introduced a review research work that tackled the utilized techniques, used for breast cancer detection using in mammography samples. Different types of neural models were reviewed, such as the hybrid adaptation in breast cancer detection. In addition, numerous artificial neural networks were utilized for detecting and diagnosing the breast cancer in [7-9]. The presented approaches were used for enhancing the micro-calcification based on illumination and non-regularity. The authors allocated the infected areas using iterative selection of threshold level method. This was done by rebuilding the shape of images and removing the redundant pixels. In addition, the introduced approaches extracted the features of these images for detecting the breast cancer. The obtained results expressed the high accuracy of performance in comparison with the previous approaches. The same approach was adopted by authors of the research work of [10-17] that were focused on the deep learning techniques. The authors of [18-23] tackled the problem of applying the software engineering technology in cooperation with the deep learning technology. Most of the previous work consider the Glopal Positioning System (GPS) and web applications to finalize the outcome productions, particularly in allocation terms [24-26].

This paper presents a CNN based deep-learning model for building an early breast cancer detection system. The proposed system uses the digital mammography images after applying the pre-processing stage. The proposed algorithm of detecting the breast cancer based on the changes of soft tissues is built based on software engineering process model. This model tackles the problem of reliability, flexibility and extendibility of the designed approach. It is important to note that the proposed system adopts a website design for easing the access of the system in the country side places. This system is designed for these places as they suffer from lack in specialist medical staff. Therefore, the system can detect the disease in early stages from the images and referring the patient to the central hospitals for providing the required treatments. It also can reduce the load on the central hospital by limitation the number of referring cases. The obtained results show the efficiency of the proposed system in terms of accuracy up to 90%, reducing the load on central hospitals and saving life of patients.

2. PROPOSED SYSTEM

The proposed system is based on designing an electronic site for detecting the breast cancer at the early stages. The system is managed by professional General Purposes (GP) health centers at the country sides of countries. This is due to the lack in specialist doctors as well as reduce the waiting queue for patients at the central breast cancer hospitals. This section is divided into numerous subsections for easing the reading flow.

2.1. System block diagram

Figure 1 illustrates the general block diagram of the proposed system. This figure explains the working steps of the proposed system in terms of user and professional registration and feeding the patent

images to the system website. The server side of the system performing the designed deep-learning model for diagnosing the breast cancer. The entered image is diagnosed to infected or non-infected, in which the patient either discharged from the GP or refereeing to central hospital.

Figure 2 shows the designing model of the proposed deep-learning based breast cancer. The convolutional neural network (CNN) is adopted for processing the matching and preparing the training dataset. The system is designed based on two classes; infected and non-infected. The classes and appointed labels for receiving data is entered to the deep-learning model. In addition, the training images are fed to the same model for performing the training model using CNN. The outcome trained model is used for diagnosing the test images into infected or non-infected.
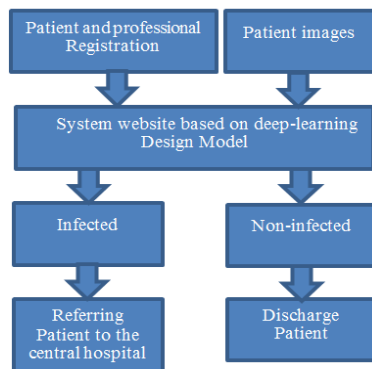
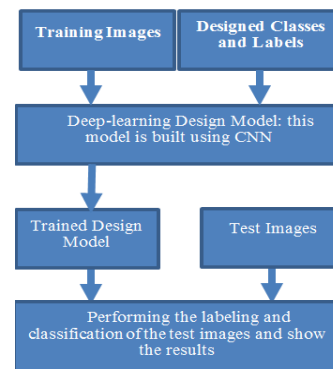Figure 1. General block diagram of the proposed system

Figure 2. Block diagram of the proposed deep-learning model

2.2. Designed software engineering process model

The software engineering process model is adopted in designing the proposed algorithms of the presented system. The reason behind using the technique of software engineering is for increasing the reliability of the proposed system and taking to the consideration any future developments. These developments include the expandability and flexibility in terms of increasing the size of involved GPs and number of users. Figure 3 explains the designed software engineering process model, used for constructing the proposed algorithms. It is well shown that the requirements of the proposed algorithms play as the core of designing the software engineering process model. The first phase takes care from collecting these requirements and classifies them into two main classes, which are infected and non-infected. While, the second phase designs the initial version of the proposed algorithm considering the requirements. There is a feedback between the phase one and two for confirming that the initial design is done according to the required limitations. The third phase develops the designed algorithm in its initial step to recover any drawback appeared throughout the completion process. The final version of the designed algorithm is implemented using the deep-learning method. The implantation is evaluated by testing the proposed algorithm over different case studies of the dataset that includes images of patients.

2.3. The proposed deep-learning algorithm

It is well known that the deep-learning model is based on building a trained dataset using the training model and diagnosing the test images using test model.

2.3.1. Training model

The training model uses the proposed trained algorithm that can be summarized as steps flow:
− Appointing the adopted classes and labeling the data.
− Classifying the training dataset.
− Extracted the adopted features from training dataset.
− Constructing the graph of CNN method.
− Checking the validity of classes and labeled dataset.
− Doing the preprocessing operations on the training images.
− Detecting any possible distortion for applying the distortions processes.
− Evaluating the 'bottleneck' image for possible saving.
− Creating the required processing layers.

- Evaluate the accuracy of the created layer.
- Setting up the required weights to their initial default values.
- Constructing the required features in iteration method.
- Locating the input bottleneck values by frequemtly freshing with distortions, or fetching from cache.
- Running a training steps.
- Capturing training summaries for TensorBoard with the `merged` operation.
- Operating the validation step and storing intermediate results.
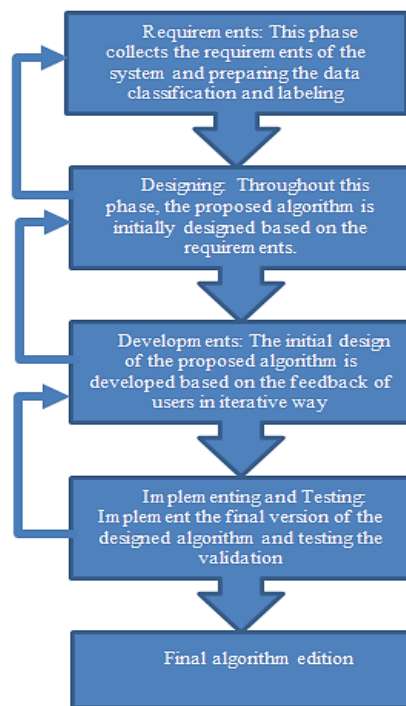- Writing out the trained and labels dataset.

Figure 3. Designed software engineering process model

2.3.2. Testing model

The obtained training dataset from the trained model, the tested images are entered to the system for diagnosing. It is drawn as steps flow:
- Entering the image file
- Feeding the image into the loaded graph as input of it.
- Obtaining the prediction set to show labels of first prediction in order of confidence.
- Obtaining the results.

2.4. The proposed GUI and algorithms

Visual Studio Code (VSC) environment is utilized to design and implement the GUI of the proposed system's web application. Figure 4 shows the home page of the proposed web application which provides a user with a useful links and information about the breast cancer. This page allows the authorized user to use the whole functions of the system after the login process done successfully.

The main process of this page is to enable the authorized users to take advantage of all system activities after the login process completes correctly. New or unregistered user can take advantage only from the information posted on this page in addition to the useful links. Figure 5 shows the flowchart of the home page. Figure 6 shows the registration page that allows the user to add new employee or new patient to the system's database from separated pages. The Registering New Employee page is shown in Figure 7. Through this page, the user must enter all the required information in the associated fields. When the submit button is pressed, a comparison process will be done between the entered information and all information stored in the employee's table of the system's database. If this employee has been registered in advance, a warning message will appear which tells user that the registration process is not done. However, if the entered information doesn't match any employee's information in the database, the registration process is done successfully.

Besides, the registering new patient page is shown in Figure 8. At this page, the registration process is carried out in the same manner as the registering a new employee process by considering the differences between the required information. As well as additional information are required which are: the genetic information. This information is composed the genetic history of the disease in the patient's family which are taken under consideration during the diagnosis process. Figure 9 shows the flowchart of the registration processes.
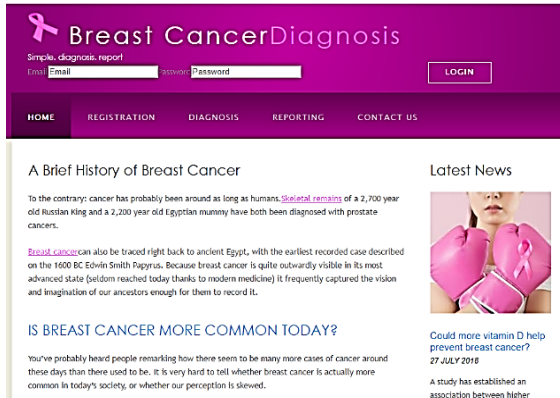
Figure 4. Home page
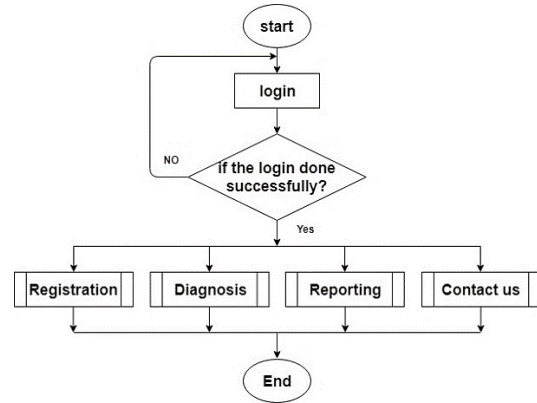
Figure 5. Flowchart of the home page process
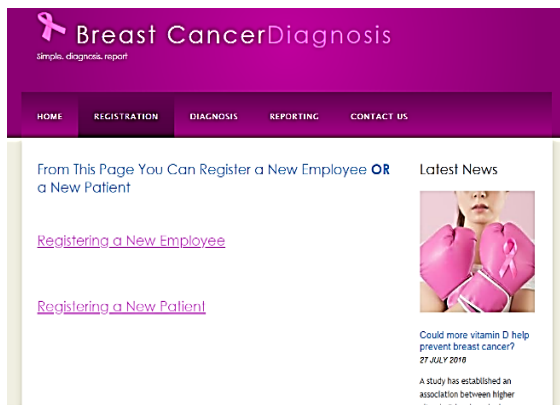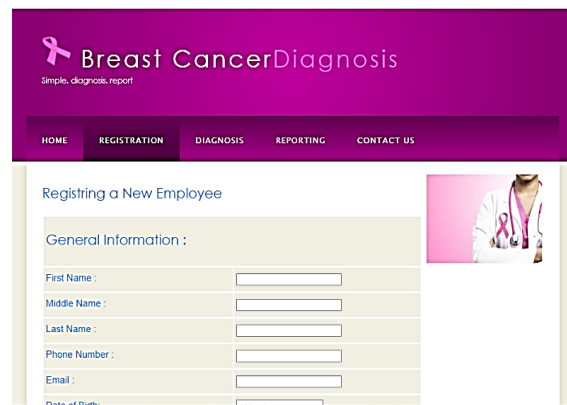

Figure 6. Registration page

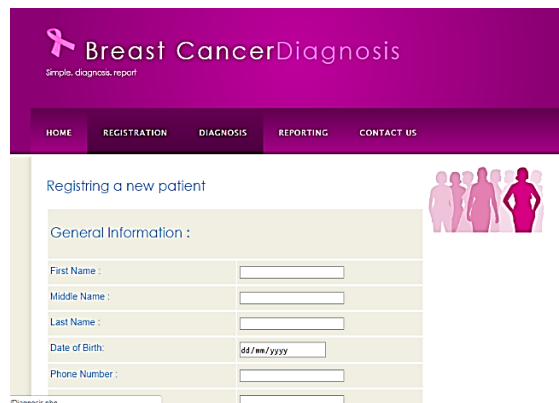Figure 7. Registering new employee page

Figure 8. Registering new patient page

Figure 10 shows the diagnosis page which composes the main function of our system. From this page, the user selects the patient (who is already stored in the database) then the mammography breast x-ray for this patient

will uploaded to the system to be manipulated using our model. The model outputs the result which will either infected or uninfected. If the uninfected result is shown, the risk factor for this patient will be calculated. Otherwise, if the infected result is shown, a drop-down list will appear that allows user to select any pre-registered hospital that the result will send to Figure 11 shows the flowchart of the diagnosis process. While, Figure 12 shows the reporting page that provide the user with a whole information about the employees, patients and infected and uninfected patients. This process is done when user clicks on the corresponding button as shown in the Figure 13. Moreover, a contact page was provided in our system in order to allow user for sending any message to us using his email as shown in Figure 14. To complete this process, the user must fill his name and a correct email at the corresponding fields as required. When the message is sent successfully, a confirm message will be sent to the sender. Figure 15 shows the flowchart of the contact us process.
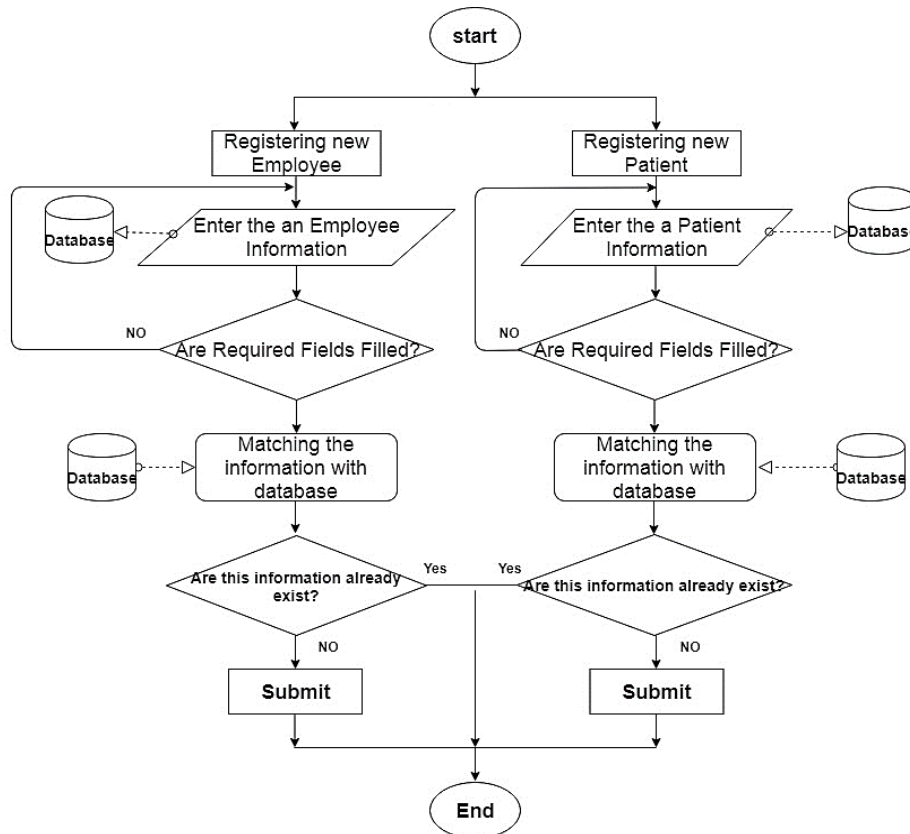

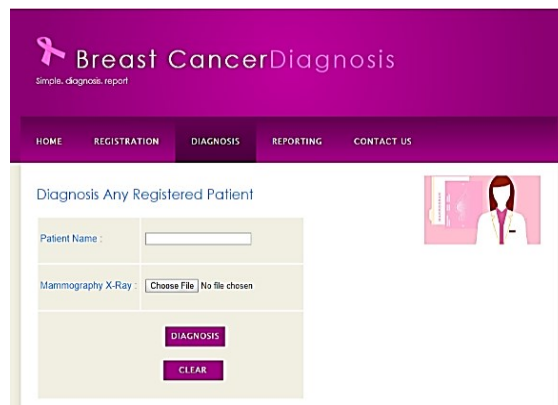Figure 9. Flowchart of the registration processes
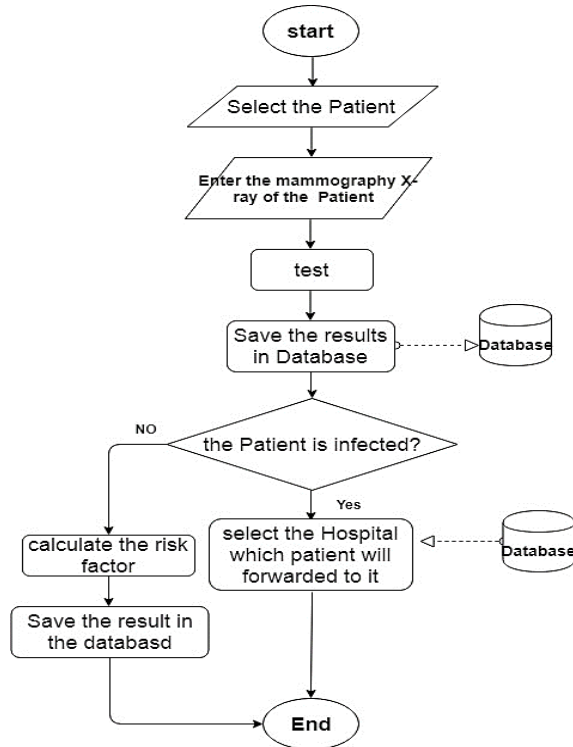

Figure 10. Diagnosis page

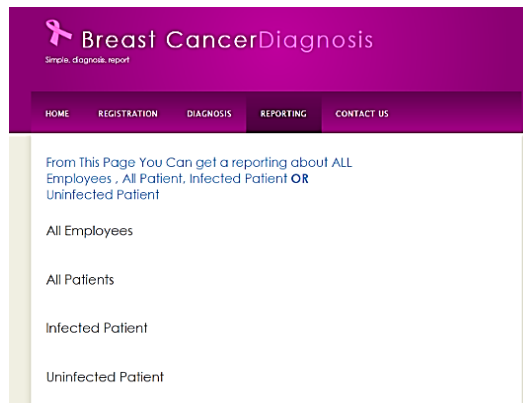Figure 11. Flowchart of the diagnosis process

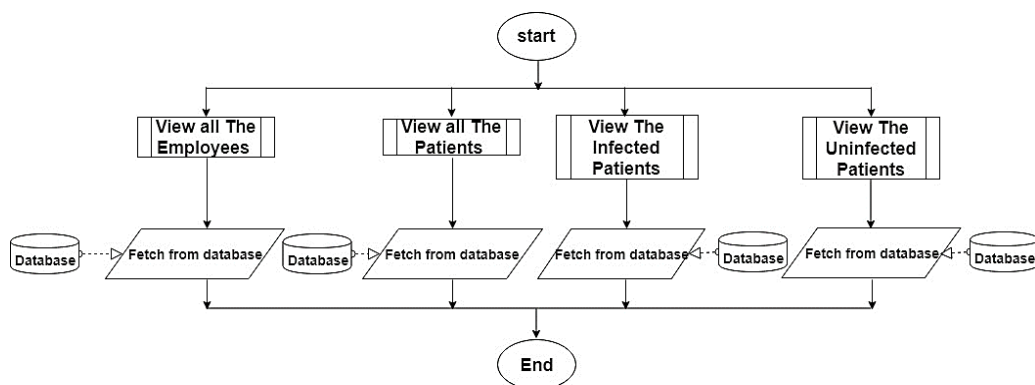Figure 12. Reporting page

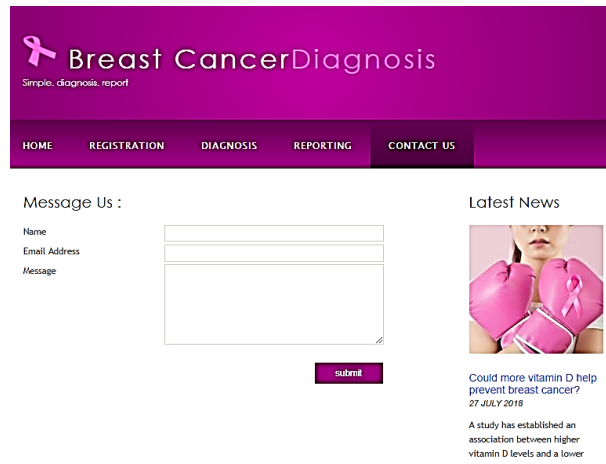Figure 13. Flowchart of the reporting process
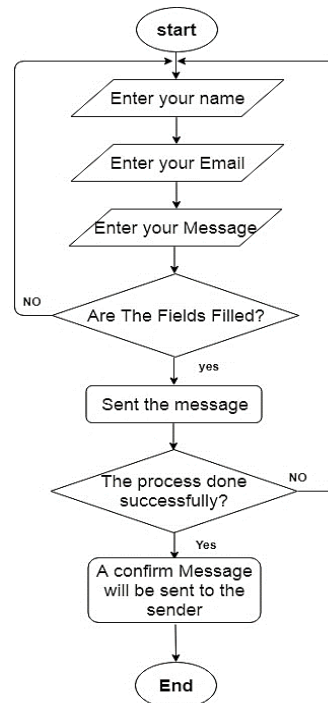
Figure 14. Contact us page

Figure 15. Flowchart of the contact us process

3.   EXPERMENTAL RESULTS AND ANALYSIS

The proposed system is tested over data set of 500 images of mammography types. Table 1 explains the classification of the utilized dataset based on the ratios of each category. The testing set category represents 30% of the total dataset. While, 70% of the dataset is allocated as training dataset. The results are obtained using HP laptop with 2.4 GHz processor, 4GB RAM supported with dedicated display adapter of (2 GB) and under operating system of Windows 10 pro. With these specifications, the proposed method is run in efficient way with processing time up to half hour from the initial point. The results are divided into two parts: deep-learning and website. The deep-learning results explain the performance of the proposed algorithm with the adopted dataset. While, the website results show the behavior of the proposed system with the testing cases that requires from the system to diagnose the infection.

3.1. Deep-learning results

Figure 16 shows the computed accuracy of training process. This accuracy is calculated from the entered training dataset. It is viewed from this figure that the accuracy is improved with the increasing of

adopted steps. This is because of the enlarging in the stored dataset of training and features from CNN. When the system crosses the step of 2000, the accuracy reaches a ratio of 100%. It is concluded from this figure that the training process achieves the acceptable ratio of the required accuracy. The accuracy is adopted as important factor for evaluating the efficiency of the proposed deep-learning algorithm. Moreover, the pre-processing operations, performed in terms of image processing, enhance the initial images to be ready for feature extraction in CNN trained model. This is to increase the efficiency of the proposed system as the noised and blurry images can affect the performance, harshly.

In order to test the validation accuracy of the proposed method, Figure 17 describes this validation as a result of detecting the breast cancer of the testing dataset. This figure proves the high validation of the results of the proposed method in training and testing phases. Figure 9 shows the validation accuracy of almost 90% at the processing step 2000 and over. It is highlighted from this figure that the accuracy is varied from 50% at the lower processing step and reached up to 90% over step 2000. The validation accuracy is being in the acceptable level after step 2000 for the same reasons of increasing the training accuracy and reducing the cross entropy. As a result of the testing outcome, the proposed method proves its efficiency in terms of training accuracy, cross entropy and validation accuracy. Although, the collected dataset is not prepared for computer programming use, the preprocessing functions performed by the proposed method reduces these effects to very minimum value of error ratio.

Table 1. The classifications of dataset

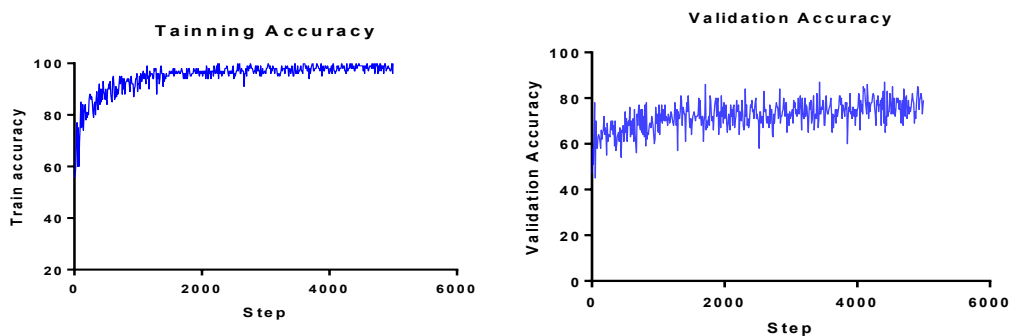| Dataset	No	Classification		Ratio
		Infected	Non-Infected	
Testing set	150	70	80	30%
Training set	350	200	150	70%
Total dataset	500	270	230	100%

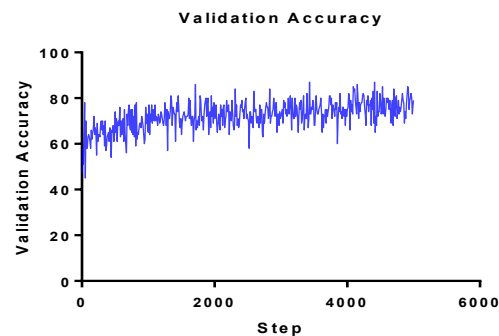Figure 16. Training accuracy

Figure 17. The computed validation accuracy

3.2. Website results

Throughout the system operating, Figure 18 explains the test results of the whole system as a website representation. In this figure, the test sample, which is the mammography image, has been submitted to the proposed system and the obtained results show the patient is infected. These results are achieved by selecting the diagnosis button. Normally, the proposed system referees the patient to the special health center at the big hospital for next step of treatments. The CLEAR button is used for erasing the results and looking for the next case study.

At the other side, Figure 19 shows the system results of uninfected case. In this figure, the mammography image of a patient is submitted to the proposed system for testing and diagnosing. The obtained results show that the patient is not infected, but the other factor is the risk. The risk factor expresses the probability of infection for patients. It can be evaluated as:

$$RiskFactor = (0.4 \times R_{mother}) + (0.3 \times R_{Grand}) + (\frac{0.3}{N_{sister}} \times R_{sister}) \tag{1}$$

where,

R_{mother} is the mother infection, it is either 1 or 0.
R_{Grand} is the grandmother infection, it is either 1 or 0.

R_{sister} is the sister infection, it is either 1 or 0.

N_{sister} is the number of infected sisters.

It is important to note that the risk factor is an indication to monitor the still not infected patient with inherited cases including mother, grandmother and sisters. We give a high ratio risk to the infected mother and less for others. This is for inheritance reasons.

Figure 18. Website results of infected patient

Figure 19. Website results of uninfected patient and risk factor

4.   CONCLUSION

In this paper, an early detection of breast cancer system based on mammography images was proposed. The proposed algorithm was formulated depending on the software engineering model to grantee the scalability, flexibility and reliability. The deep-learning technology has been utilized for detecting the changes in the soft tissues at the investigated mammography images. The proposed system adopted a website for GUI design. The website allowed the doctors and patients to access the system regardless the distances and places. At the other hand, the proposed system considered the computing of risk factor of uninfected patients. This risk factor offered a monitoring indicator for patients under risk. The proposed system was tested in two categories. The first one tested the accuracy of the designed deep-learning algorithm. While the other one considered whole system testing representing as website results. The obtained results showed the efficiency of the proposed system in terms of accuracy and early detection of breast cancer.

REFERENCES

[1]   Saraswathi Duraisamy, Srinivasan Emperumal, "Computer-aided mammogram diagnosis system using deep learning convolutional fully complex-valued relaxation neural network classifier," *IET Computer Vision*, vol. 11, no. 8, pp. 656-662, 2017.

[2]   Michiel Kallenberg, Kersten Petersen, Mads Nielsen, Andrew Y. Ng., Pengfei Diao, Christian Igel, Celine M. Vachon, Katharina Holland, Rikke Rass Winkel, Nico Karssemeijer, and Martin Lillholm, "Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1322-1331, 2018.

[3]   Gustavo Carneiro, Jacinto Nascimento, and Andrew P. Bradley, "Automated Analysis of Unregistered Multi-View Mammograms with Deep Learning," *IEEE Transections on Medical Imaging*, vol. 36, no. 11, pp. 2355-2365, 2017.

[4]   Moi Hoon Yap, Gerard Pons, Joan Mart´ı, Sergi Ganau, Melcior Sent´ıs, Reyer Zwiggelaar, Adrian K. Davison, and Robert Mart´, "Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 4, pp. 1218-1226, 2018.

[5]   Xiaofei Zhang, Yi Zhang, Erik Y. Han, Nathan Jacobs, Qiong Han, Xiaoqin Wang, Jinze Liu, "Classification of Whole Mammogram and Tomosynthesis Images Using Deep Convolutional Neural Networks," *IEEE Transactions on NanoBioscience*, vol. 17, no. 3, pp. 237-242, 2018.

[6]   M. M. Mehdy, P. Y. Ng., E. F. Shair, N. I. Md Saleh, and C. Gomes, "Artificial Neural Networks in Image Processing for Early Detection of Breast Cancer," *Computational and Mathematical Methods in Medicine*, vol. 2017, pp. 1-15, 2017.

[7]   Vikas Kumar Saubhagyal, Asha Rani and Vijander Singh, "ANN based Detection of Breast Cancer in Mammograph Images," *1st IEEE International Conference on Power Electronics. Intelligent Control and Energy Systems*, 2016.

[8]   Ghada Saad, Ahmad Khadour, Qosai Kanafani, "ANN and Adaboost Application for Automatic Detection of Microcalcifications in breast cancer," *The Egyptian Journal of Radiology and Nuclear Medicine*, vol. 47, no. 4, pp. 1803-14, 2016.

[9]   Dominik Eckert, Sulaiman Vesal, Ludwig Ritschl, Steffen Kappler, and Andreas Maier, "Deep Learning-Based Denoising of Mammographic Images Using Physics-Driven Data Augmentation," arXiv:1912.05240v, 2019.

[10] Wei Guo, Ran Wu, Yanhua Chen and Xinyan Zhu, "Deep Learning Scene Recognition Method Based on Localization Enhancement," *Sensors*, vol. 18, no. 10, pp. 1-20, 2018.

[11] François Chollet, "Deep Learning with Python," *Manning,* 2018.

[12] Martin Thoma, "Analysis and Optimization of Convolutional Neural Network Architectures," Master Thesis in University of the State of Baden-Wuerttemberg and National Research Center of the Helmholtz Association, 2017.

[13] Yunus Çelik, Mahmut Altun, and Mahit Güneş, "Color based moving object tracking with an active camera using motion information," *IEEE Artificial Intelligence and Data Processing Symposium (IDAP),* pp. 1-4, 2017.

[14] Puja Prasad, Rashmi Pathak, Vinit Kumar Gunjan, H. V. Ramana Rao, "Deep Learning Based Representation for Face Recognition," *ICCCE,* pp. 419-424, 2019.

[15] Nguyen L.D., Lin D., Lin Z., Cao J., "Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation," *In Proceedings of the 2018 IEEE International Symposium on Circuitsand Systems (ISCAS),* pp. 1-5, 2018.

[16] Xuyu Wang, Lingjun Gao and Shiwen Mao, "BiLoc: Bi-Modal Deep Learning for Indoor Localization with Commodity 5GHz WiFi," *IEEE access*, vol. 5, pp. 4209-4220, 2017.

[17] Yuan Y., Mou L., Lu X., "Scene recognition by manifold regularized deep learning architecture," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 26, no. 10, pp. 222-233, 2015.

[18] Ian Sommerville, "Software Engineering," *10th Edition, Pearson Education, Inc,* 2017.

[19] Miikka Kuutilaa, Mika Mantylaa, Umar Farooqa, and Maelick Claes, "Time Pressure in Software Engineering: A Systematic Review," *Informayion and Softare Technology*, vol. 121, 2020.

[20] L. Mottola, G. P. Picco , F. J. Oppermann, "Simplifying the Integration of Wireless Sensor Networks into Business Processes," *IEEE Transaction on Software Engineering*, vol. 45, no. 6, pp. 576-596, 2019.

[21] Ayuba J, Safwana H, and Abdulazeez Y, "Software Development of Integrated Wireless Sensor Networks for RealTime Monitoring of Oil and Gas Flow Rate Metering Infrastructure," *Journal of Information Technology & Software Engineering*, vol. 8, no. 2, pp. 1-10, 2018.

[22] Iman Almomani, and Afnan Alromi , "Integrating Software Engineering Processes in the Development of Efficient Intrusion Detection Systems in Wireless Sensor Networks," *Sensors*, vol. 20, no. 5, pp. 1-28, 2020.

[23] Asmaa Munem, and Muayad Croock, "Smart Traffic Light Control System for Emergency Ambulance," *International Journal of Advanced Research in Computer Engineering*, vol. 5, no. 8, pp. 2247-55, 2016.

[24] Evizal Abdul Kadir, Hitoshi Irie, Sri Listia Rosa, and Mahmod Othman, "Modelling of wireless sensor networks for detection land and forest fire hotspot," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 17, no. 6, pp. 2772-2781, 2019.

[25] Shawn M. Boeser, "Global Positioning Systems (GPS)," Encyclopedia of Coastal Science, 2019.

[26] Jordi Mas and Carles Matue, "Introduction to Web Application Development," *Eureca Media,* 2015.